

(19) 世界知的所有権機関  
国際事務局



(43) 国際公開日  
2001 年 5 月 31 日 (31.05.2001)

PCT

(10) 国際公開番号  
WO 01/39175 A1

(51) 国際特許分類: G10L 11/02,  
21/02, H04B 1/64, H04J 3/17, H04Q 7/32

(SUZUKI, Kaori) [JP/JP]. 大田 恭士 (OTA, Yasuji)  
[JP/JP]; 〒211-8588 神奈川県川崎市中原区上小田中4  
丁目1番1号 富士通株式会社内 Kanagawa (JP).

(21) 国際出願番号: PCT/JP99/06539

(22) 国際出願日: 1999 年 11 月 24 日 (24.11.1999)

(74) 代理人: 古谷史旺(FURUYA, Fumio); 〒160-0023 東京  
都新宿区西新宿1丁目19番5号 第2明宝ビル9階 Tokyo  
(JP).

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(81) 指定国 (国内): JP, US.

(71) 出願人 (米国を除く全ての指定国について): 富士通  
株式会社 (FUJITSU LIMITED) [JP/JP]; 〒211-8588 神  
奈川県川崎市中原区上小田中4丁目1番1号 Kanagawa  
(JP).

添付公開書類:  
— 国際調査報告書

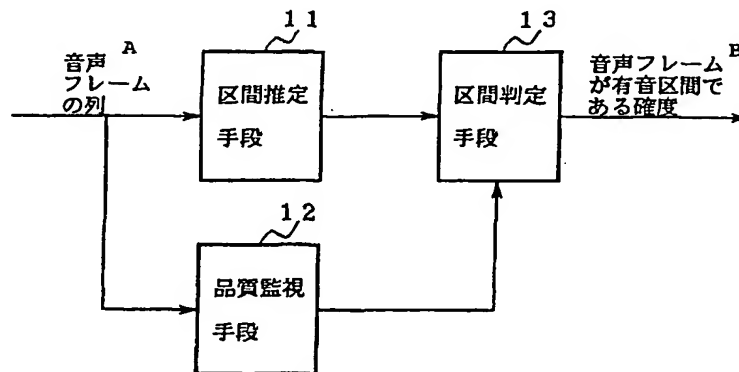
(72) 発明者; および

(75) 発明者/出願人 (米国についてのみ): 鈴木香緒里

2 文字コード及び他の略語については、定期発行される  
各 PCT ガゼットの巻頭に掲載されている「コードと略語  
のガイダンスノート」を参照。

(54) Title: METHOD AND APPARATUS FOR VOICE DETECTION

(54) 発明の名称: 音声検出装置および音声検出方法



A ... SERIES OF VOICE FRAMES

B ... CONFIDENCE OF VOICE FRAME WITHIN VOICED INTERVAL

11 ... INTERVAL ESTIMATE MEANS

12 ... QUALITY MONITOR MEANS

13 ... INTERVAL DETERMINATION MEANS

(57) Abstract: A voice-detecting device is provided to precisely discriminate between voiced and unvoiced intervals adaptively to various features of possible noises superimposed on a voice signal. The voice detection device comprises interval estimate means (11) for determining the probability that each of the time series voice frames belongs to a voiced interval according to a statistical technique; quality monitor means (12) for monitoring the quality of the voice signal of each voice frame; and interval determination means (13) for weighting the resulting probability based on the signal quality to determine the confidence of a voice interval.

[続葉有]

WO 01/39175 A1

(19) 日本国特許庁 (JP)

# 再公表特許 (A 1)

(11) 国際公開番号

WO 01/039175

発行日 平成15年5月27日 (2003. 5. 27)

(43) 国際公開日 平成13年5月31日 (2001. 5. 31)

(51) Int. Cl.<sup>7</sup>

識別記号

F I

G 1 0 L 11/02

G 1 0 L 9/00

D

21/02

H 0 4 B 1/64

H 0 4 B 1/64

H 0 4 J 3/17

H 0 4 Q 7/32

G 1 0 L 9/00

F

H 0 4 J 3/17

H 0 4 B 7/26

V

審査請求 未請求 予備審査請求

有

(全 4 6 頁)

出願番号 特願2001-540759 (P2001-540759)

(21) 国際出願番号 PCT/JP99/06539

(22) 国際出願日 平成11年11月24日 (1999. 11. 24)

(81) 指定国 J P, U S

(71) 出願人 富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番1号

(72) 発明者 鈴木 香緒里

神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

(72) 発明者 大田 恭士

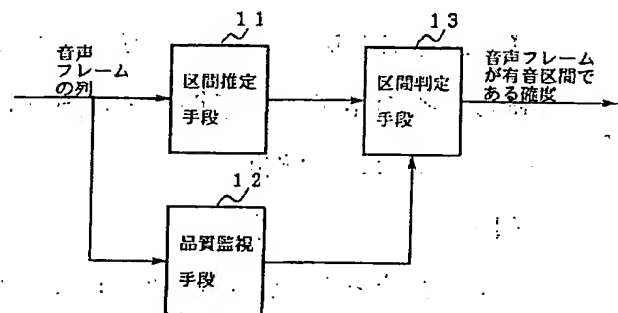
神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

(74) 代理人 弁理士 古谷 史旺

(54) 【発明の名称】 音声検出装置および音声検出方法

(57) 【要約】

本発明は、音声検出装置および音声検出方法に関する。  
本発明の目的は、音声信号に重畳され得る雑音の多様な特徴に柔軟に適用して確度高く有音区間と無音区間とを識別することである。そのために、本発明にかかわる音声検出装置は、時系列の順に与えられる音声フレーム毎に、音声信号の統計的な性質に基づいて有音区間に属する確率を求める区間推定手段 1 1 と、音声フレーム毎に音声信号の品質を監視する品質監視手段 1 2 と、個々のフレームについて、求められた確率に上記の品質を重み付け、有音区間である確度を得る区間判定手段 1 3 とを備えて構成される。



**【特許請求の範囲】**

**【請求項1】** 音声信号として時系列の順に与えられる音声フレーム毎に、その音声信号の統計的な性質に基づいて有音区間に属する確率を求める区間推定手段と

前記音声フレーム毎に前記音声信号の品質を監視する品質監視手段と、

前記音声信号として時系列の順に与えられる個々のフレームについて、前記区間推定手段によって求められた確率に、前記品質監視手段によって監視された品質を重み付け、前記有音区間である確度を得る区間判定手段と

を備えたことを特徴とする音声検出装置。

**【請求項2】** 音声信号として時系列の順に与えられる音声フレーム毎に、その音声信号の統計的な性質に基づいて有音区間に属する確度を求める区間判定手段と

前記音声フレーム毎に前記音声信号の品質を監視する品質監視手段とを備え、

前記区間判定手段は、

前記音声フレーム毎に、個別に含まれる前記音声信号の瞬時値の列に、前記品質監視手段によって監視された品質の単調減少関数あるいは単調非増加関数として与えられる重みによる重み付けを行う

ことを特徴とする音声検出装置。

**【請求項3】** 時系列の順に与えられる個々の音声フレームについて、圧伸処理を施し、その結果として得られた音声信号の瞬時値の列をその音声信号の統計的な性質に基づいて解析することによって、有音区間に属する確度を求める区間判定手段と、

前記音声フレーム毎に前記音声信号の品質を監視する品質監視手段とを備え、

前記区間判定手段は、

前記音声フレーム毎に、前記品質監視手段によって監視された品質の単調減少関数として与えられる圧伸特性を前記圧伸処理に適用する

ことを特徴とする音声検出装置。

**【請求項4】** 請求の範囲1に記載の音声検出装置において、

品質監視手段は、

音声信号の有音区間と無音区間との双方または何れか一方についてその音声信号の特徴を求め、これらの特徴の何れか一方もしくは両者の差分としてこの音声信号の品質を得る

ことを特徴とする音声検出装置。

【請求項 5】請求の範囲 2 に記載の音声検出装置において、

品質監視手段は、

音声信号の有音区間と無音区間との双方または何れか一方についてその音声信号の特徴を求め、これらの特徴の何れか一方もしくは両者の差分としてこの音声信号の品質を得る

ことを特徴とする音声検出装置。

【請求項 6】請求の範囲 3 に記載の音声検出装置において、

品質監視手段は、

音声信号の有音区間と無音区間との双方または何れか一方についてその音声信号の特徴を求め、これらの特徴の何れか一方もしくは両者の差分としてこの音声信号の品質を得る

ことを特徴とする音声検出装置。

【請求項 7】請求の範囲 1 に記載の音声検出装置において、

品質監視手段は、

音声フレーム毎に、雑音推定パワーを求め、その雑音推定パワーの単調非増加関数として音声信号の品質を得る

ことを特徴とする音声検出装置。

【請求項 8】請求の範囲 2 に記載の音声検出装置において、

品質監視手段は、

音声フレーム毎に、雑音推定パワーを求め、その雑音推定パワーの単調非増加関数として音声信号の品質を得る

ことを特徴とする音声検出装置。

【請求項 9】請求の範囲 3 に記載の音声検出装置において、

品質監視手段は、

音声フレーム毎に、雑音推定パワーを求め、その雑音推定パワーの単調非増加

関数として音声信号の品質を得る

ことを特徴とする音声検出装置。

【請求項 10】請求の範囲 1 に記載の音声検出装置において、

品質監視手段は、

音声フレーム毎に、雑音推定パワーと SN 比の推定値とを求め、前者の単調非増加関数と、後者の単調非減少関数として音声信号の品質を得る

ことを特徴とする音声検出装置。

【請求項 11】請求の範囲 2 に記載の音声検出装置において、

品質監視手段は、

音声フレーム毎に、雑音推定パワーと SN 比の推定値とを求め、前者の単調非増加関数と、後者の単調非減少関数として音声信号の品質を得る

ことを特徴とする音声検出装置。

【請求項 12】請求の範囲 3 に記載の音声検出装置において、

品質監視手段は、

音声フレーム毎に、雑音推定パワーと SN 比の推定値とを求め、前者の単調非増加関数と、後者の単調非減少関数として音声信号の品質を得る

ことを特徴とする音声検出装置。

【請求項 13】請求の範囲 1 に記載の音声検出装置において、

品質監視手段は、

音声フレーム毎に、標準化確率変数を求め、その標準化確率変数の単調減少関数として音声信号の品質を得る

ことを特徴とする音声検出装置。

【請求項 14】請求の範囲 2 に記載の音声検出装置において、

品質監視手段は、

音声フレーム毎に、標準化確率変数を求め、その標準化確率変数の単調減少関数として音声信号の品質を得る

ことを特徴とする音声検出装置。

【請求項 15】請求の範囲 3 に記載の音声検出装置において、

品質監視手段は、

音声フレーム毎に、標準化確率変数を求め、その標準化確率変数の単調減少関数として音声信号の品質を得る

ことを特徴とする音声検出装置。

【請求項 16】 請求の範囲 1 に記載の音声検出装置において、

品質監視手段は、

音声フレーム毎に、標準化確率変数と SN 比の推定値とを求め、前者の単調非増加関数と、後者の単調非減少関数として音声信号の品質を得る

ことを特徴とする音声検出装置。

【請求項 17】 請求の範囲 2 に記載の音声検出装置において、

品質監視手段は、

音声フレーム毎に、標準化確率変数と SN 比の推定値とを求め、前者の単調非増加関数と、後者の単調非減少関数として音声信号の品質を得る

ことを特徴とする音声検出装置。

【請求項 18】 請求の範囲 3 に記載の音声検出装置において、

品質監視手段は、

音声フレーム毎に、標準化確率変数と SN 比の推定値とを求め、前者の単調非増加関数と、後者の単調非減少関数として音声信号の品質を得る

ことを特徴とする音声検出装置。

【請求項 19】 音声信号として時系列の順に与えられる音声フレーム毎に、その音声信号の統計的な性質に基づいて有音区間に属する確率を求め、

前記音声フレーム毎に前記音声信号の品質を監視し、

前記音声信号として時系列の順に与えられる個々のフレームについて、前記求められた確率に、前記監視された品質を重み付け、前記有音区間である確度を得る

ことを特徴とする音声検出方法。

【請求項 20】 音声信号として時系列の順に与えられる音声フレーム毎に、その音声信号の統計的な性質に基づいて有音区間に属する確度を求め、

前記音声フレーム毎に前記音声信号の品質を監視し、

前記音声フレーム毎に、個別に含まれる前記音声信号の瞬時値の列に、前記監

視された品質の単調減少関数あるいは単調非増加関数として与えられる重みによる重み付けを行う

ことを特徴とする音声検出方法。

【請求項 21】 時系列の順に与えられる個々の音声フレームについて、圧伸処理を施し、その結果として得られた音声信号の瞬時値の列をその音声信号の統計的な性質に基づいて解析することによって、有音区間に属する確度を求め、

前記音声フレーム毎に前記音声信号の品質を監視し、

前記音声フレーム毎に、前記監視された品質の単調減少関数として与えられる圧伸特性を前記圧伸処理に適用する

ことを特徴とする音声検出方法。

**【発明の詳細な説明】****技術分野**

本発明は、音声信号を取り込み、その音声信号の有音区間と無音区間とを識別する音声検出装置と、その音声検出装置に適用された音声検出方法に関する。

**背景技術**

近年、デジタル信号処理技術が高度に進展し、かつ移動通信システムその他の通信システムにおいては、これらのデジタル信号処理技術が適用されることによって、伝送情報である音声信号に実時間で多様な信号処理が施されている。

また、このような通信システムの送信端では、伝送帯域の圧縮や無線周波数の有効利用に併せて、消費電力の節減を目的として、上述した音声信号の無音区間と有音区間とを検出し、この有音区間に限って伝送路に対する送信を許容する音声検出装置が搭載されている。

図12は、音声検出装置が搭載された無線端末装置の構成例を示す図である。

図において、マイク41は音声検出装置42の入力と送受信部43の変調入力とに接続され、その送受信部43のアンテナ端子にはアンテナ44の給電端が接続される。音声検出装置42の出力は送受信部43の送信制御入力に接続され、その送受信部43の制御用の入出力には制御部45の対応する入出力ポートが接続される。制御部45の特定の出力ポートは音声検出装置42の制御入力に接続され、かつ送受信部43の復調出力はレシーバ46の入力に接続される。

このような構成の無線端末装置では、送受信部43はマイク41とレシーバ46とを介して送受されるべき伝送情報である音声信号と、アンテナ44を介してアクセスが可能な無線伝送路（図示されない。）との無線インタフェースをとる。

制御部45は、その送受信部43と連携することによって、この無線伝送路の形成に要するチャネル制御を主導的に行う。

音声検出装置42は、所定の周期で上述した音声信号をサンプリングすることによって音声フレームの列を生成する。さらに、音声検出装置42は、これらの個々の音声フレームについて、音声信号の性質に基づいて、有音区間と無音区間との何れに該当するかの識別を行い、その識別の結果を示す2値信号を出力する



。なお、上述した性質とは、例えば、下記の事項である。

- ・ 約55デシベルのダイナミックレンジを有する。
- ・ 振幅分布が標準的な確率密度関数で近似され得る。
- ・ エネルギーの密度と零交差回数とは、それぞれ無音区間と有音区間とにおいて異なる値となる。

送受信部43は、その2値信号の論理値が上述した無音区間を意味するする期間には、送信を見合わせる。

すなわち、音声信号に伝送情報として有効な情報が含まれていない期間には、送受信部43が無用に送信することが規制される。したがって、消費電力の節減にあわせて、他の無線チャネルに介する干渉の抑制と無線周波数の有効利用とが図られる。

しかし、このような従来例では、マイク41を介して与えられる音声信号に大きなレベルの雑音が重畳している期間には、有音区間と無音区間との間における特徴量（例えば、上述した零交差回数）の差が小さな値となる。

さらに、有音区間であっても子音区間には、一般に、音声信号の振幅は、母音区間に比べて小さい値に多く分布する。

したがって、子音区間は無音区間として識別される可能性が高く、上述したように誤って識別された子音（有音）区間には、該当する音声フレームの送信が行われず、通話品質が無用に劣化する可能性が高かった。

また、上述した雑音のレベルが過大である場合には、その騒音が重畳された音声信号の大半を示す音声フレームに亘って送信が規制される可能性があった。

なお、これらの問題点については、例えば、有音区間としての識別がされ易い値にその識別の基準となる特徴量等の閾値が設定されることによって、改善され得る。

しかし、このような閾値が適用された場合には、無音区間であるにもかかわらず有音区間として識別される確率が増加し、かつその有音区間の時間率が過大となり得るために、既述の消費電力の節減、干渉の抑制および無線周波数の有効利用が十分に図られない可能性があった。

### 発明の開示

本発明の目的は、音声信号とその音声信号に重畳され得る雑音との多様な特徴に柔軟に適応し、かつ確度高く有音区間と無音区間とを識別できる音声検出装置および音声検出方法を提供することである。

また、本発明の目的は、有音区間の内、例えば、子音区間のように音声信号の振幅が小さい領域に多く分布する区間について、音声信号の品質が低い場合であっても有音区間である確度が大きな値で得られる点にある。

さらに、本発明の目的は、精度よく、音声フレーム毎に有音区間である確度が得られる点にある。

また、本発明の目的は、所要する処理量が削減され、あるいは応答性が高められる点にある。

さらに、本発明の目的は、重畳している雑音のレベルが大きく、かつSN比が小さい音声フレームについても、有音区間に属することを示す確度が大きな値で得られる点にある。

上述した目的は、音声フレーム毎に、有音区間に属する確率と品質とを求め、その確率をこの品質で重み付けて出力する点に特徴がある音声検出装置および音声検出方法によって達成される。

このような構成の音声検出装置および音声検出方法では、個々の音声フレームは、品質が良好であるほど、高い確率で有音区間として識別され、反対に無音区間として識別される確率が小さくなる。

また、上述した目的は、音声フレーム毎に、有音区間に属する確率と品質とを求め、この確率が求められるべき音声フレームのレベルをその品質が大きいほど小さな値に設定する点に特徴がある音声検出装置および音声検出方法によって達成される。

このような構成の音声検出装置および音声検出方法では、品質が低いほど、個々の音声フレームに含まれる個々の音声信号の瞬時値に対して大きな重み付けが行われるために、その結果である瞬時値の列として与えられる音声信号が上述した有音区間に属する確度が高い値で求められる。

さらに、上述した目的は、音声フレーム毎に、有音区間に属する確率と品質と

が求められ、その品質が高いほど、この確率が求められるべき音声フレームの圧伸処理に適用されるべき圧伸特性の勾配あるいは閾値を大きな値に設定する点に特徴がある音声検出装置および音声検出方法によって達成される。

このような構成の音声検出装置および音声検出方法では、音声信号の品質が低いほど個々の音声フレームに含まれる瞬時値に対して大きな重み付けを行う処理は、圧伸処理として行われる。

また、上述した目的は、音声フレーム毎に、有音区間と無音区間との双方あるいは何れか一方について特徴を求め、これらの特徴を品質として適用する点に特徴がある音声検出装置によって達成される。

このような構成の音声検出装置では、音声信号の品質は、音響分析あるいは音声分析を実現する多様な技術の適用の下で安定に得られる。

さらに、上述した目的は、音声フレーム毎に、雑音推定パワーを求め、その雑音推定パワーを品質として適用する点に特徴がある音声検出装置および音声検出方法によって達成される。

このような構成の音声検出装置では、上述した雑音推定パワーの算出は、一般に、単純な算術演算によって達成される。

また、上述した目的は、音声フレーム毎に、雑音推定パワーと、SN比の推定値とを求め、前者に対する単調非増加関数と後者に対する単調非減少関数として与えられる数を品質として適用する点に特徴がある音声検出装置によって達成される。

このような構成の音声検出装置では、重畳している雑音のレベルが大きく、かつSN比が小さい音声フレームについても、有音区間に属することを示す確度が大きな値で得られる。

さらに、上述した目的は、雑音推定パワーに代えて標準化確率変数が適用される点で先行して記載された音声検出装置と異なる音声検出装置によって達成される。

このような構成の音声検出装置では、標準化確率変数は、その絶対値が大きいほど、「音声フレームの振幅の先頭値が音声信号の標準的な振幅に比べて大きく、かつ大きなレベルの雑音がこの音声フレームに重畳されている可能性が高いこ

と」を意味し、反対に小さいほど、「音声フレームの振幅の先頭値が音声信号の標準的な振幅に比べて小さく、この音声フレームに重畳されている雑音のレベルも小さいこと」を意味する。

したがって、標準化確率変数は、上述した雑音推定パワーに代替し得る。

また、上述した目的は、標準化確率変数が音声フレームの振幅分布とその振幅分布の最大値とに基づいて近似的に算出される点に特徴がある音声検出装置によって達成される。

このような構成の音声検出装置では、上述した標準化確率変数は、簡便な算術演算によって求められる。

さらに、上述した目的は、音声フレームの単位に先行して得られた品質が時系列の順に積分され、その結果が品質として適用される点に特徴がある音声検出装置によって達成される。

このような構成の音声検出装置では、時系列の順に得られた音声信号の品質に伴い得る急峻な変動の成分が軽減され、あるいは抑圧される。

また、上述した目的は、音声フレームの単位に先行して得られた品質が時系列の順に積分され、その結果が大きいほどこの結果が小さく重み付けされて得られた値が品質として適用される点に特徴がある音声検出装置によって達成される。

このような構成の音声検出装置では、先行して与えられた音声フレームの品質が高く、あるいはその品質が高かった時間率が大きいほど、後続して与えられる音声フレームが有音区間である確度が大きな値で得られる。

#### 発明を実施するための最良の形態

まず、図1および図2を参照して本発明にかかわる音声検出装置および音声検出方法の原理を説明する。

図1は、本発明の第1の原理ブロック図である。

図1に示す音声検出装置は、区間推定手段11と、品質監視手段12と、区間判定手段13とから構成される。

本発明にかかわる第1の音声検出装置の原理は、下記の通りである。

区間推定手段11は、音声信号として時系列の順に与えられる音声フレーム毎に、その音声信号の統計的な性質に基づいて有音区間に属する確率を求める。ま

た、品質監視手段 12 は、音声フレーム毎に音声信号の品質を監視する。

区間判定手段 13 は、上述したように音声信号として時系列の順に与えられる個々のフレームについて、区間推定手段 11 によって求められた確率に、品質監視手段 12 によって監視された品質を重み付け、有音区間である確度を得る。

このような音声検出装置では、個々の音声フレームは、音声信号の品質が良好であるほど、高い確率で有音区間として識別され、反対に無音区間として識別される確率が小さくなる。

したがって、有音区間の内、例えば、子音区間のように音声信号の振幅が小さな領域に多く分布する区間については、その子音区間における音声信号の品質が低い場合であっても有音区間である確度が大きな値で得られる。

図 2 は、本発明の第 2 の原理ブロック図である。

図 2 に示す音声検出装置は、区間判定手段 15、15A と、品質監視手段 16 とから構成される。

本発明にかかわる第 2 の音声検出装置の原理は、下記の通りである。

区間判別手段 15 は、音声信号として時系列の順に与えられる音声フレーム毎に、その音声信号の統計的な性質に基づいて有音区間に属する確度を求める。品質監視手段 16 は、これらの音声フレームについて、個別に上述した音声信号の品質を監視する。

また、区間判定手段 15 は、上述した音声フレーム毎に、個別に含まれる音声信号の瞬時値の列に、品質監視手段 16 によって監視された品質に対して単調減少関数あるいは単調非増加関数として与えられる重みによる重み付けを行う。

このような音声検出装置では、音声フレーム毎に、品質が低いほど、区間判別手段 15 は、その音声フレームに含まれる個々の音声信号の瞬時値に大きな重み付けを行い、その結果として与えられる瞬時値の列として与えられる音声信号について、上述した有音区間に属する確度を求める。

したがって、有音区間の内、例えば、子音区間のように音声信号の振幅が小さな領域に多く分布する区間については、その子音区間における音声信号の品質が低い場合であっても有音区間である確度が大きな値で得られる。

本発明にかかわる第 3 の音声検出装置の原理は、下記の通りである。

品質監視手段 16 は、時系列の順に音声フレームの列として与えられる音声信号について、その音声フレーム毎に品質を監視する。

区間判定手段 15A は、これらの音声フレームに個別に圧伸処理を施し、その結果として得られた音声信号の瞬時値の列をその音声信号の統計的な性質に基づいて解析することによって、有音区間に属する確度を求める。

さらに、区間判定手段 15A は、上述した音声フレーム毎に、品質監視手段 16 によって監視された品質に対して音声の瞬時値の単調減少関数として与えられる圧伸特性を前記圧伸処理に適用する。

このような音声検出装置では、音声信号の品質が低いほど、個々の音声フレームに含まれる音声信号の瞬時値に対して大きな重み付けを行う処理は、上述した圧伸処理として、既述の第 2 の音声検出装置と同様に行われる。

したがって、有音区間の内、例えば、子音区間のように音声信号の振幅が小さな領域に多く分布する区間については、その子音区間における音声信号の品質が低い場合であっても、有音区間であることを示す確度が大きな値で得られる。

本発明にかかわる第 4 の音声検出装置の原理は、下記の通りである。

品質監視手段 12、16 は、音声信号の有音区間と無音区間との双方または何れか一方についてその音声信号の特徴を求め、これらの特徴の何れか一方もしくは両者の差分としてこの音声信号の品質を得る。

このような音声検出装置では、音声信号の品質は、音響分析あるいは音声分析を実現する多様な技術の適用の下で、上述した特徴として安定に得られる。

したがって、既述の第 1 ないし第 3 の音声検出装置に比べて、音声フレーム毎に有音区間である確度が精度よく得られる。

本発明にかかわる第 5 の音声検出装置の原理は、下記の通りである。

品質監視手段 12、16 は、音声フレーム毎に、雑音推定パワーを求め、その雑音推定パワーの単調減少関数として音声信号の品質を得る。

このような音声検出装置では、雑音推定パワーの算出は、一般に、単純な算術演算によって達成される。

したがって、既述の第 1 ないし第 3 の音声検出装置に比べて、処理量が削減され、あるいは応答性が高められる。

本発明にかかわる第6の音声検出装置の原理は、下記の通りである。

品質監視手段12、16は、音声フレーム毎に、雑音推定パワーとSN比の推定値とを求め、前者の単調非増加関数と、かつ後者の単調非減少関数として音声信号の品質を得る。

このような音声検出装置では、大きなレベルの雑音が重畳し、かつSN比が小さい音声フレームについても、有音区間に属することを示す確度が大きな値で得られる。

本発明にかかわる第7の音声検出装置の原理は、下記の通りである。

品質監視手段12、16は、音声フレーム毎に、標準化確率変数を求め、その標準化確率変数の単調非増加関数として音声信号の品質を得る。

このような音声検出装置では、標準化確率変数は、その絶対値が大きいほど、「音声フレームの振幅の先頭値が音声信号の標準的な振幅に比べて大きく、かつ大きなレベルの雑音がこの音声フレームに重畳されている可能性が高いこと」を意味し、反対に小さいほど、「音声フレームの振幅の先頭値が音声信号の標準的な振幅に比べて小さく、この音声フレームに重畳されている雑音のレベルも小さいこと」を意味する。

したがって、既述の第6の音声検出装置と同様に、重畳している雑音のレベルが高く、かつSN比が小さい音声フレームについても、有音区間に属することを示す確度が大きな値で得られる。

本発明にかかわる第8の音声検出装置の原理は、下記の通りである。

品質監視手段12、16は、音声フレーム毎に、標準化確率変数とSN比の推定値とを求め、前者の単調非増加関数と、かつ後者の単調非減少関数として音声信号の品質を得る。

このような音声検出装置では、大きなレベルの雑音が重畳し、かつSN比が小さい音声フレームについても、有音区間に属することを示す確度が大きな値で得られる。

本発明にかかわる第1の音声検出方法の原理は、下記の通りである。

第1の音声検出方法では、音声信号として時系列の順に与えられる音声フレーム毎に、その音声信号の統計的な性質に基づいて有音区間に属する確率が求めら

れ、かつこの音声信号の品質が監視される。

さらに、音声信号として時系列の順に与えられる個々のフレームについては、上述したように求められた確率に、監視された品質が重み付けられることによって、有音区間である確度が得られる。

このような音声検出方法では、個々の音声フレームは、音声信号の品質が良好であるほど、高い確率で有音区間として識別され、反対に無音区間として識別される確率が小さくなる。

したがって、有音区間の内、例えば、子音区間のように音声信号の振幅が小さな領域に多く分布する区間については、その子音区間における音声信号の品質が低い場合であっても有音区間である確度が大きな値で得られる。

本発明にかかわる第2の音声検出方法の原理は、下記の通りである。

第2の音声検出方法では、音声信号として時系列の順に与えられる音声フレーム毎に、その音声信号の統計的な性質に基づいて有音区間に属する確率が求められ、かつこの音声信号の品質が前記音声フレーム毎に監視される。

さらに、音声フレーム毎に、個別に含まれる音声信号の瞬時値の列については、上述したように監視された品質に対して単調減少関数あるいは単調非増加関数として与えられる重みによる重み付けが行われる。

このような音声検出方法では、音声フレーム毎に、音声信号の品質が低いほど、その音声フレームに含まれる個々の音声信号の瞬時値に対して大きな重み付けが行われ、その結果として与えられる瞬時値の列として与えられる音声信号について、上述した有音区間に属する確度が求められる。

したがって、有音区間の内、例えば、子音区間のように音声信号の振幅が小さな領域に多く分布する区間については、その子音区間における音声信号の品質が低い場合であっても有音区間である確度が大きな値で得られる。

本発明にかかわる第3の音声検出方法の原理は、下記の通りである。

第3の音声検出方法では、時系列の順に与えられる個々の音声フレームについて、圧伸処理が施され、その結果として得られた音声信号の瞬時値の列がその音声信号の統計的な性質に基づいて解析されることによって、有音区間に属する確度が求められ、かつその音声信号の品質が監視される。



さらに、上述した圧伸処理の過程では、音声フレーム毎に、このようにして監視された品質の単調減少関数として与えられる圧伸特性が適用される。

このような音声検出方法では、音声信号の品質が低いほど、個々の音声フレームに含まれる音声信号の瞬時値に対して大きな重み付けを行う処理は、上述した圧伸処理として、第二の音声検出方法と同様に行われる。

したがって、有音区間の内、例えば、子音区間のように音声信号の振幅が小さな領域に多く分布する区間については、その子音区間における音声信号の品質が低い場合であっても有音区間である確度が大きな値で得られる。

以下、図面に基づいて本発明の実施形態について詳細に説明する。

図3は、本発明の実施形態1、3～8を示す図である。

図において、図12に示すものと機能および構成が同じものについては、同じ符号を付与して示し、ここでは、その説明を省略する。

本実施形態と図12に示す従来例との構成の相違点は、音声検出装置42に代えて音声検出装置20が備えられた点にある。

音声検出装置20は、初段に備えられた有音／無音識別部21と、その有音／無音識別部21のモニタ出力に直結されたモニタ端子を有し、この有音／無音識別部21と共に初段に備えられた識別確度判定部22と、これらの有音／無音識別部21および識別確度判定部22の出力にそれぞれ接続された2つのポートを有するメモリ23と、そのメモリ23の出力に直結されると共に、最終段として備えられた最終判定部24とから構成される。

図4は、実施形態1の動作フローチャートである。

#### [実施形態1]

以下、図3および図4を参照して本発明にかかわる実施形態1の動作を説明する。

音声検出装置20では、有音／無音識別部21は、マイク41を介して与えられる音声信号に、図12に示す音声検出装置42が行う処理と同じ処理を施すことによって、既述の音声フレーム毎に有音区間と無音区間との識別を行い、かつメモリ23と識別確度判定部22とに並行してこの識別の結果を示す2値情報I<sub>t</sub>を与える。

なお、この2値情報 $I_t$ の論理値については、簡単のため、有音区間には「1」に設定され、反対に無音区間には「0」に設定されると仮定する。

一方、識別確度判定部22は、有音／無音識別部21と並行して上述した音声信号を既述の音声フレームの列に変換する。さらに、識別確度判定部22は、有音／無音識別部21によって与えられる2値情報 $I_t$ の論理値に応じて有音区間と無音区間とを識別し、これらの区間について、個々の音声フレームの特徴量 $F_t$ （ここでは、簡単のため、エネルギーと零交差回数との双方あるいは何れか一方であると仮定する。）の分布（平均値）を定常的に監視する。

さらに、識別確度判定部22は、個々の音声フレームが与えられる期間に、有音区間と無音区間とにおける上述した特徴量 $F_t$ の分布（平均値）の差分が所定の閾値 $F_{th}$ を下回るか否かの判別を行い、その判別の結果を示す2値の識別確度 $R_t$ を求める。

なお、このような識別確度 $R_t$ の論理値については、上述した差分が閾値 $F_{th}$ を下回る程度に音声信号の品質が低い場合には「0」に設定され、反対にこの閾値 $F_{th}$ を上回る程度に音声信号の品質が良好である場合には「1」に設定されると仮定する。

また、メモリ23には、有音／無音識別部21によって与えられた2値情報 $I_t$ と、識別確度判定部22によって求められた識別確度 $R_t$ とが上述した音声フレームの単位に対応付けられて蓄積される。

最終判定部24は、このようにメモリ23に蓄積された2値情報 $I_t$ と識別確度 $R_t$ との個々の組み合わせに応じて、下記の処理を順次行う。

- ・ 識別確度 $R_t$ の論理値が「1」である場合には、論理値が2値情報 $I_t$ の論理値に等しい2値信号を送受信部43に与える（図4（1））。

- ・ 識別確度 $R_t$ の論理値が「0」である場合には、論理値が「1」である2値信号を送受信部43に与える（図4（2））。

また、送受信部43は、制御部45が行うチャネル制御の下で割り付けられた無線チャネルに、マイク41によって与えられる音声信号で変調された送信波信号を送信する過程には、上述したように、音声フレーム単位に音声検出装置20で行われる処理の演算所要時間に等しい遅延を与えることによって、この音声検

出装置 20 との同期を維持する。

すなわち、音声信号の品質が良好である場合には、有音／無音識別部 21 によって与えられた 2 値情報  $I_t$  が 2 値信号として送受信部 43 に与えられるが、その品質が良好でない場合には、この 2 値信号の論理値は、有音区間を示す「1」に設定される。

したがって、本実施形態によれば、識別確度  $R_t$  の如何にかかわらず音声信号の統計的な性質のみに基づいて有音区間と無音区間との識別が行われる従来例に比べて、その音声信号の品質が悪い有音区間が無音区間として識別されることが確度高く回避され、かつ伝送品質の劣化が緩和される。

なお、本実施形態では、有音／無音識別部 21 と識別確度判定部 22 とが並行して個別に音声信号を音声フレームの列に変換する処理を行っている。

しかし、このような処理は、有音／無音識別部 21 と識別確度判定部 22 との何れか一方によって主導的に行われ、あるいはこれらの有音／無音識別部 21 と識別確度判定部 22 との前段に配置された手段によって行われてもよい。

また、本実施形態では、有音／無音識別部 21 によって求められた 2 値情報  $I_t$  と、識別確度判定部 22 によって求められた識別確度  $R_t$  とが音声フレーム毎に対応つけられてメモリ 23 に蓄積されている。

しかし、メモリ 23 は、有音／無音識別部 21、識別確度判定部 22 および最終判定部 24 によって行われるべき既述の処理の所要時間に伴い得る変動分が許容される程度に小さい場合には、備えられなくてもよい。

さらに、本実施形態では、送受信部 43 は、音声フレーム単位に音声検出装置 20 によって行われる処理の演算所要時間に等しい遅延を与えることによって、この音声検出装置 20 との同期を維持している。

しかし、このような遅延は、上述した同期が所望の確度で維持される程度に小さい場合には、何ら与えられなくてもよい。

また、本実施形態では、識別確度判定部 22 によって既述の識別確度  $R_t$  が求められている。

しかし、識別確度判定部 22 と最終判定部 24 とは、例えば、その識別確度判定部 22 が下記の処理の何れかのみを行うことによって、如何なる形態で機能分

散が図られてもよい。

- ・ 既述の音声フレームが与えられる時点あるいは期間に、有音区間と無音区間における上述した特徴量  $F_t$  の分布（平均値）を求める。

- ・ その特徴量  $F_t$  の分布（平均値）を求め、この隔たり（差分）が所定の閾値  $F_{th}$  を下回るか否かの判別を行う。

さらに、本実施形態では、有音区間と無音区間との特徴量  $F_t$  の差分と閾値  $F_{th}$  との大小関係に基づいて、音声信号の品質の良否が判別されている。

しかし、本発明は、このような構成に限定されず、例えば、有音区間と無音区間との何れか一方の特徴量が所望の確度で既知の値として与えられる場合には、他方の特徴量のみが求められ、その特徴量と規定の閾値との大小関係に基づいて、音声信号の伝送品質の良否が判別されてもよい。

#### [実施形態2]

図5は、本発明の実施形態2を示す図である。

図において、図3に示すものと機能および構成が同じものについては、同じ符号を付与して示し、ここでは、その説明を省略する。

本実施形態と既述の実施形態1との構成の相違点は、音声検出装置20に代えて音声検出装置30が備えられた点にある。

音声検出装置30と音声検出装置20との構成の相違点は、有音／無音識別部21に代えて有音／無音識別部21Aが備えられ、最終判定部24に代えて識別条件調整部31が備えられ、その識別条件調整部31の出力は送受信部43の対応する制御入力に代わる有音／無音識別部21Aの閾値入力に接続され、その制御入力に有音／無音識別部21Aの出力が接続された点にある。

図6は、実施形態2の動作フローチャートである。

以下、図5および図6を参照して本発明にかかわる実施形態2の動作を説明する。

本実施形態と実施形態1との相違点は、識別条件調整部31が行う下記の処理と、有音／無音識別部21Aがその処理の下で与えられる閾値に基づいて既述の2値情報  $I_t$  を求める点とにある。

なお、以下では、有音／無音識別部21A、識別確度判定部22およびメモリ

23が連係することによって行われる処理の手順については、既述の実施形態1と基本的に同じであるので、ここでは、その説明を省略する。

有音／無音識別部21Aは、マイク41を介して与えられる音声信号に、図12に示す従来例に搭載された音声検出装置42が行う処理と同じ処理を施し、その処理の過程でこの音声信号の統計的な性質にかかわる閾値（以下、「区間識別閾値」という。）として、識別条件調整部31によって与えられる値を適用することによって、2値情報 $I_t$ を求める。

また、識別条件調整部31は、このようにして求められた2値情報 $I_t$ と識別確度判定部22によって求められた識別確度 $R_t$ との組み合わせを順次メモリ23を介して取り込み、かつ下記の処理を行う。

- ・ 識別確度 $R_t$ の論理値が「1」である場合には、その有音／無音識別部21に、『上述した音声信号の品質が良好である期間に2値情報 $I_t$ を求める過程で、有音／無音識別部21Aが適用すべき標準的な区間識別閾値（以下、「標準閾値」という。）』を与える（図6（1））。なお、標準閾値については、識別条件調整部31に予め与えられていると仮定する。

- ・ 識別確度 $R_t$ の論理値が「0」である場合には、先行して有音／無音識別部21Aに与えられた区間識別閾値（上述した「標準閾値」であってもよい。）を下記の何れかの値に更新し、あるいは設定する（図6（2））。

- －有音／無音識別部21Aが後続する音声フレームを有音区間に属する音声フレームと識別する可能性が高い値

- －有音／無音識別部21Aが後続する音声フレームを有音区間に属する音声フレームとして確実に識別する値

さらに、送受信部43は、有音／無音識別部21Aによって与えられる2値情報 $I_t$ の列を既述の2値信号として取り込み、既述の実施形態1と同様に音声検出装置30との同期を維持する。

このように本実施形態によれば、音声信号の品質が良好である場合には、有音／無音識別部21Aによって与えられた2値情報 $I_t$ が2値信号として送受信部43に与えられるが、その品質が良好でない場合には、区間識別閾値が適宜更新されることによって、『有音区間を示す「1」にこの2値信号の論理値が設定さ

れる確率』が高められる。

したがって、本実施形態によれば、識別確度  $R_t$  の如何にかかわらず音声信号の統計的な性質のみに基づいて有音区間と無音区間との識別が行われる従来例に比べて、品質が悪い有音区間が無音区間として識別されることに起因する伝送品質の劣化が緩和され、あるいは回避される。

なお、本実施形態では、区間識別閾値は、識別条件調整部 31 によって適宜更新され、あるいは設定されている。

しかし、本発明は、このような構成に限定されず、例えば、有音／無音識別部 21A に音声信号を線形領域で増幅する可変利得増幅器が搭載され、かつ有音区間と無音区間との識別の基準がその音声信号のレベルである場合には、上述した区間識別閾値に代えてこの可変利得増幅器の利得が可変されてもよい。

### 〔実施形態 3〕

本実施形態と実施形態 1 との構成の相違点は、識別確度判定部 22 に代えて識別確度判定部 22A が備えられた点にある。

図 7 は、実施形態 3 の動作フローチャートである。

以下、図 3 および図 7 を参照して本実施形態の動作を説明する。

本実施形態の特徴は、識別確度判定部 22A が行う下記の処理の手順にある。

識別確度判定部 22A は、有音／無音識別部 21 と並行して音声信号を音声フレームの列に変換し（図 7（1））、かつ個々の音声フレームについて下記の処理を施す。

なお、以下では、個々の音声フレームは、簡単のため、時系列  $t$  ( $= 0 \sim N$ ) の順に  $(N+1)$  個の瞬時値  $x(t)$  の列として与えられると仮定する。

1. 下式（1）で示される算術演算を行うことによって、フレームパワー  $P_t$  を算出し、時系列  $t$  の順に蓄積する（図 7（2））。
2. 先行する音声フレームについて同様に算出され、かつ蓄積されている先行フレームパワー  $P_{t-i}$  を取得する（図 7（3））。
3. 規定の時定数  $\alpha$  ( $< 1$ ) に対して下式（2）で示される算術演算を行うことによって、雑音推定パワー  $P_{Nt}$  を指数平滑法に基づいて算出する（図 7（4））。

4. この雑音推定パワー $P_{Nt}$ と、その雑音推定パワー $P_{Nt}$ について既述の閾値 $F_{th}$ と同様に予め設定された閾値 $P_{th}$ とを比較することによって、前者が後者を上回るか否かの判別を行い（図7（5））、その判別の結果を示す2値の識別確度 $R_t$ を求める（図7（6））。

$$P_{Nt} = \alpha P_{t-1} + (1 - \alpha) P_t \quad \cdots (2)$$

$$P_t = \sum_{n=0}^N x(n)^2 \quad \cdots (1)$$

なお、このような識別確度 $R_t$ の論理値については、上述した判別の結果が真である場合には「0」（通話信号の品質が低いことを意味する。）に設定され、反対に偽である場合には「1」（通話信号の品質が良好であることを意味する。）に設定されると仮定する。

また、最終判定部24は、既述の実施形態1と同様にこの識別確度 $R_t$ を参照することによって2値信号を生成し、その2値信号を送受信部43に順次与える。

このように本実施形態によれば、上式（1）、（2）に示す単純な算術演算によって通話信号の品質が簡便に求められ、かつ上述した判別の結果が偽である期間は、有音／無音識別部21によって与えられた2値情報の論理値 $I_t$ の如何にかかわらず、高い確度で、あるいは確実に有音期間として識別される。

〔実施形態4〕

本実施形態と実施形態1との構成の相違点は、識別確度判定部22に代えて識別確度判定部22Bが備えられた点にある。

図8は、実施形態4の動作フローチャートである。

以下、図3および図8を参照して本実施形態の動作を説明する。

本実施形態の特徴は、識別確度判定部22Bが行う下記の処理の手順にある。

識別確度判定部22Bは、有音／無音識別部21と並行して音声信号を音声フレームの列に変換し（図8（1））、かつ個々の音声フレームについて下記の処理を施す。

1. 既述の実施形態3において識別確度判定部22Aによって行われる処理の手順と同じ手順に基づいてフレームパワー $P_t$ と、雑音推定パワー $P_{Nt}$ とを算

出する（図 8（2））。

2. 下式（3）で示される算術演算を行うことによって、この音声フレームの SN 比の推定値（以下、単に、「SN 推定値」という。） $SN_t$  を算出する（図 8（3））。

3. この SN 推定値  $SN_t$  がその SN 推定値  $SN_t$  について既述の閾値  $F_{th}$  と同様に予め設定された閾値  $SN_{th}$  を上回るか否かの判別（以下、「SN 判別」という。）を行う（図 8（4））。

4. 上述した雑音推定パワー  $P_{N_t}$  が既述の閾値  $P_{th}$  を下回るか否かの判別（以下、「雑音判別」という。）を行う（図 8（5））。

5. これらの判別の結果の組み合わせに応じて、下記の通りに識別確度  $R_t$  を求め、かつ出力する。

① SN 判別の結果が真である場合と、その SN 判別の結果が偽であり、かつ雑音判別の結果が真である場合とには、この雑音判別の結果を示す 2 値を識別確度  $R_t$  として出力する（図 8（6））。

② SN 判別の結果が偽であり、かつ雑音判別の結果が偽である場合には、論理値が「0」である識別確度  $R_t$  を出力する（図 8（7））。

$$SN_t = 10 \log_{10}(P_t / P_{N_t}) \quad \dots (3)$$

すなわち、SN 推定値  $SN_t$  が小さく、かつ上述した雑音推定パワー  $P_{N_t}$  が大きい場合に、有音／無音識別部 21 によって行われた識別の確度が著しく低下した状態であっても、最終判定部 74 によって有音区間が無音区間と識別されることが確度高く回避される。

#### 〔実施形態 5〕

本実施形態と実施形態 1 との構成の相違点は、識別確度判定部 22 に代えて識別確度判定部 22C が備えられた点にある。

図 9 は、実施形態 5 の動作フローチャートである。

以下、図 3 および図 9 を参照して本実施形態の動作を説明する。

本実施形態と既述の実施形態 4 との相違点は、識別確度判定部 22C が行う下記の処理の手順にある。



識別確度判定部 22C は、有音／無音識別部 21 と並行して音声信号を音声フレームの列に変換し（図 9（1））、かつ個々の音声フレームについて、雑音推定パワー  $P_{N_t}$  を算出する処理に代えて以下の処理を行う。

- A) 時系列  $t$  の順に与えられる個々の音声フレームで示される音声信号の振幅の先頭値  $s_{P_t}$  と平均値  $s_{m_t}$  とを求めて蓄積する。
- B) 最新の音声フレームが与えられる度に、所定の数  $M$  に対してその音声フレームに先行する時点で時系列  $t$  の順にそれぞれ与えられた  $M$  個の音声フレームについて、同様に蓄積された先頭値  $s_{P_t}$  と平均値  $s_{m_t}$  とを取得する。
- C) これらの先頭値および平均値を下式（4）に代入することによって行われる算術演算の結果として、該当する音声フレームで示される音声信号の振幅の標準偏差  $\sigma_t$  を算出する。
- D) 最新の音声フレームで示される音声信号の振幅の先頭値  $x$  を求める。
- E) これらの標準偏差  $\sigma_t$  と先頭値  $x$  とに対して下式（5）で示される算術演算を行うことによって、上述した音声信号の振幅の標準化確率変数  $P_{r_t}$  を算出する（図 9（2））。

$$\sigma_t = \left[ \left\{ \sum_{i=t-M}^{t-1} (s_{P_i} - s_{m_i})^2 / M \right\} \right]^{1/2} \quad \cdots (4)$$

$$P_{r_t} = x / \sigma_t \quad \cdots (5)$$

なお、標準化確率変数  $P_{r_t}$  は、最新の音声フレームに含まれる音声信号の振幅の先頭値  $s_{P_t}$  と、その振幅の分布との相関関係を意味する。

さらに、標準化確率変数  $P_{r_t}$  は、その絶対値が大きいほど、「最新の音声フレームの振幅の先頭値が音声信号の標準的な振幅に比べて大きく、かつ大きなレベルの雑音がこの音声フレームに重畳されている可能性が高いこと」を意味し、反対に小さいほど、「最新の音声フレームの振幅の先頭値が音声信号の標準的な振幅に比べて小さく、この音声フレームに重畳されている雑音のレベルが小さいこと」を意味する。

また、識別確度判定部 22C は、実施形態 4 と同様にして  $SN$  推定値  $SN_t$  を

求め（図 9（3））、かつ「SN判定」を行う（図 9（4））。

さらに、識別確度判定部 22C は、上述した標準化確率変数  $P_r$  が規定の閾値  $P_{rth}$  を下回るか否かの判別（以下、「変数判別」という。）を行う（図 9（5））。

また、識別確度判定部 22C は、これらの判別の結果の組み合わせに応じて、下記の通りに識別確度  $R_t$  を求めて出力する。

I. SN判別の結果が真である場合と、変数判別の結果が真である場合には、この変数判別の結果を示す 2 値を識別確度  $R_t$  として出力する（図 9（6））。

II. SN判別の結果が偽であり、かつ変数判別の結果が偽である場合には、論理値が「0」である識別確度  $R_t$  を出力する（図 9（7））。

すなわち、識別確度  $R_t$  の論理値は、標準化確率変数  $P_r$  の値が大きい場合には、有音／無音識別部 21 によって行われる識別の確度が著しく低下した状態であっても、最終判定部 74 によって有音区間を無音区間と識別することが確度高く回避される。

#### [実施形態 6]

本実施形態と実施形態 5 との構成の相違点は、識別確度判定部 22 に代えて識別確度判定部 22D が備えられた点にある。

図 10 は、実施形態 6 の動作フローチャートである。

以下、図 3 および図 10 を参照して本実施形態の動作を説明する。

本実施形態と実施形態 5 との相違点は、識別確度判定部 22D が識別確度判定部 22C に代えて、後述する手順に基づいて標準化確率変数  $P_r$  を算出する点にある。

音声信号の振幅分布を示す確率密度関数は、一般に、ガンマ分布やラプラス分布で近似が可能である。

また、この確率密度関数  $P(x)$  は、例えば、上述したラプラス分布で近似された場合には、標準偏差で正規化された音声の振幅  $x$  に対して、下記の式で定義される。

$$P(x) = (1/\sqrt{2}) \exp(-\sqrt{2} \cdot |x|)$$

したがって、標準偏差で正規化された音声の振幅  $x$  の絶対値は、

$$|x| = (-1/\sqrt{2}) \cdot \ln(\sqrt{2} \cdot P(x)) \quad \dots \quad (6)$$

の式で与えられる。

ところで、個々の音声フレームに含まれ、かつサンプリングされて所定のデジタル信号処理が施される標本値の数  $K$ （ここでは、簡単のため、「1000」であると仮定する。）は、一般に、既知の値として与えられる。

また、このような場合には、個々の音声フレームに含まれる音声フレームに振幅の先頭値が出現する確率は、 $(1/K)$  で与えられる。

識別確度判定部 22D は、この確率  $(= 1/K)$  が上式 (6) に適用されることによって得られる下記の式で示される算術演算を行い、その結果結果として  $|x|$  の値を求める (図 10 (1))。

$$\begin{aligned} |x| &= (-1/\sqrt{2}) \cdot \ln(\sqrt{2} \cdot (1/K)) \\ &= (-1/\sqrt{2}) \cdot \ln(\sqrt{2} \cdot (1/1000)) \end{aligned}$$

さらに、識別確度判定部 22D は、該当する音声フレームで与えられる音声信号の振幅の瞬時値  $p$  を求め (図 10 (2))、その瞬時値  $p$  と上述した  $|x|$  の値とに対して、

$$\sigma_t = p / |x|$$

の式で示される算術演算を行うことによって標準偏差  $\sigma_t$  を算出する (図 10 (3)) と共に、この標準偏差  $\sigma_t$  の値を既述の式 (5) に代入することによって標準化確率変数  $P_{rt}$  を求める (図 10 (4))。

すなわち、実施形態 5 において行われる既述の処理 A) ~ E) に比べて簡便な算術演算に基づいて標準化確率変数  $P_{rt}$  が求められる。

したがって、本実施形態によれば、実施形態 5 に比べて所望の応答性が得られるために確保されるべき処理量の削減、あるいはその応答性の向上が可能となる。

なお、本実施形態では、識別確度判定部 22D は、単位音声フレーム毎に既述の処理を行なっている。

しかし、このような処理については、時系列の順に与えられる所望の複数の音声フレーム毎に同様の処理が行われることによって、誤差の圧縮が図られてもよい。

なお、実施形態 3～実施形態 6 は、実施形態 1 の構成に既述の変更が施されることによって構成されている。

しかし、これらの実施形態については、実施形態 2 の構成に同様の発明が適用されることによって構成されてもよい。

#### 〔実施形態 7〕

本実施形態の構成は、既述の実施形態 1～実施形態 6 の構成の何れと同じであってもよい。

図 11 は、実施形態 7 および実施形態 8 の動作フローチャートである。

以下、図 3、図 5 および図 11 を参照して、本実施形態の動作を説明する。

本実施形態の特徴は、既述の識別確度判定部 22、22A～22D の何れかによって行われる下記の処理の手順にある。

なお、以下では、簡単のため、識別確度判定部 22、22A～22D の内、識別確度判定部 22 のみに着目することとする。

識別確度判定部 22 は、新たな識別確度  $R_t$  が求められても、その識別確度  $R_t$  をメモリ 23 に直接格納せず、時系列の順に所定の重み付けを行いつつ積分することによって得られた積分値（以下、「積分識別確度  $R_{It}$ 」という。）を求め（図 11（1））、その積分識別確度  $R_{It}$  を識別確度  $R_t$  に代えてメモリに格納する（図 11（2））。

このような積分の過程では、時系列の順に求められた識別確度  $R_t$  に伴い得る急峻な変動の成分は、上述した重み付けに適用された重みに応じて軽減され、あるいは抑圧される。

したがって、本実施形態によれば、音声信号に伴い得る多様な騒音に対する柔軟な適応が可能となり、かつ実施形態 1～実施形態 6 の何れについても、本発明が適用されることによって、性能の安定化が図られる。

なお、本実施形態では、上述した重みだけではなく、積分を実現する算術演算の形態やアルゴリズムが具体的に示されていない。

しかし、このような算術演算の過程では、先行して所定の数 $C$ に亘って得られた識別確度 $R_t$ に移動平均法、指数平滑法その他の如何なるアルゴリズムおよび重みによる積分処理が行われてもよい。

#### 〔実施形態8〕

本実施形態の構成は、既述の実施形態1～7の構成と基本的に同じである。

以下、図3、図5および図11を参照して本実施形態の動作を説明する。

本実施形態の特徴は、識別確度判定部22、21A～22Dが行う下記の処理の手順にある。

本実施形態と既述の実施形態7との相違点は、識別確度判定部22、21A～22Dが下記の処理を行う点にある。

なお、以下では、簡単のため、識別確度判定部22、22A～22Dの内、識別確度判定部22のみに着目することとする。

識別確度判定部22は、新たな積分識別確度 $R_{I,t}$ が求められても、メモリ23は、その積分識別確度 $R_{I,t}$ を直接格納しない。

さらに、識別確度判定部22は、新たな積分識別確度 $R_{I,t}$ が求められると、その積分識別確度 $R_{I,t}$ を内部に備えられたレジスタ（図示されない。）に保持する（図11（a））。

また、識別確度判定部22は、この積分識別確度 $R_{I,t}$ が後述する閾値 $R_{I,t,h}$ を上回るか否かを判別し（図11（b））、その判別の結果を示す2値情報 $R_{B,t}$ をその積分識別確度 $R_{I,t}$ に代えてメモリ23に格納する（図11（c））。

さらに、識別確度判定部22は、下記の処理を行うことによって、後続して与えられる音声フレームに施される同様の処理に適用されるべき閾値 $R_{I,t,h}$ を確定する（図11（d））。

- ・ 上述したレジスタに保持された積分識別確度 $R_{I,t}$ の値が大きいほど、小さな値に設定する。
- ・ 反対に、その積分識別確度 $R_{I,t}$ の値が小さいほど、大きな値に設定する

すなわち、識別確度 $R_t$ 、積分識別確度 $R_{I_t}$ に代えて、メモリ23を介して最終判定部24あるいは識別条件調整部31に与えられるべき2値情報 $R_{B_t}$ の論理値は、先行して与えられた音声フレームの品質が高く、あるいはその品質が高かった時間率が大きいほど、後続して与えられる音声フレームが有音区間として識別される確率が高くなる値に設定される。

したがって、本実施形態によれば、実施形態1～実施形態7に比べて、有音区間が無音区間として識別されることに起因する伝送品質の低下が確度高く回避される。

なお、上述した各実施形態では、

- ・ 有音／無音識別部21、21Aが求める2値情報 $I_t$ 、
  - ・ 識別確度判定部22、22A～22Dが求める2値の識別確度 $R_t$ 、積分識別確度 $R_{I_t}$ および2値情報 $R_{I_t}$ の何れか、
  - ・ 最終判定部24によって送受信部43に与えられる2値信号の値、
- の何れもが2値情報となっている。

しかし、これらの値については、既述の目的が達成される限り、多値の情報として与えられ、かつ閾値との大小関係の判別に代えて量子化が行われ、あるいは適宜重み付けが施されてもよい。

さらに、上述した各実施形態では、無線伝送系の送信部に本発明が適用されている。

しかし、本発明は、このような無線伝送系に限定されず、有線伝送系の送信部、あるいは音声に応答して所定の処理（パターン認識を含む。）や動作を行う多様な電子機器にも同様に適用が可能である。

以下、既述の実施形態として開示された発明の内、請求の範囲1ないし請求の範囲2-1として記載された発明以外の発明を順次「追加開示請求の範囲」として列記する。

なお、下記の各「追加開示請求の範囲」の原理ブロック図は、図1および図2に示す通りである。

（追加開示請求の範囲1）

請求の範囲 7 ないし請求の範囲 1 2 の何れか 1 項に記載の音声検出装置において、

品質監視手段 1 2、1 6 は、

音声フレーム毎に、個別に含まれる音声信号の瞬時値の先頭値を求め、その音声信号の振幅分布を近似する確率密度関数に、これらの瞬時値の数と先頭値が出現する確率とを適用することによって、この確率密度関数の標準偏差で正規化された振幅を算出すると共に、その振幅と先頭値との比として標準化確率変数を求める

ことを特徴とする音声検出装置。

(追加開示請求の範囲 2)

請求の範囲 1 ないし請求の範囲 1 8 および追加開示請求の範囲 1 の何れか 1 項に記載の音声検出装置において、

品質監視手段 1 2、1 6 は、

得られた音声信号の品質を順次積分し、その結果を正規の品質として適用することを特徴とする音声検出装置。

(追加開示請求の範囲 3)

請求の範囲 1 ないし請求の範囲 1 8 および追加開示請求の範囲 1、2 の何れか 1 項に記載の音声検出装置において、

品質監視手段 1 2、1 6 は、

得られた音声信号の品質を順次積分し、その結果の単調増加関数あるいは単調非減少関数として得られる値をこの品質として適用する

ことを特徴とする音声検出装置。

以下、追加開示請求の範囲 1～3 の作用・効果を順次記述する。

追加開示請求の範囲 1 にかかわる音声検出装置では、品質監視手段 1 2、1 6 は、音声フレーム毎に、個別に含まれる音声信号の瞬時値の先頭値を求め、その音声信号の振幅分布を近似する確率密度関数に、これらの瞬時値の数と先頭値が出現する確率とを適用することによって、この確率密度関数の標準偏差で正規化された振幅を算出すると共に、その振幅と先頭値との比として標準化確率変数を求める。

このような音声検出装置では、上述した標準化確率変数は、既述の第5の音声検出装置に比べて、簡便な算術演算に基づいて求められる。

したがって、上述した第5の音声検出装置に比べて、所望の応答性が得られるために確保されるべき処理量の削減、あるいはその応答性の向上が可能となる。

追加開示請求の範囲2にかかわる音声検出装置では、品質監視手段12、16は、得られた音声信号の品質を順次積分し、その結果を正規の品質として適用する。

このような音声検出装置では、時系列の順に得られた音声信号の品質に伴い得る急峻な変動の成分が軽減され、あるいは抑圧される。

したがって、本発明にかかわる音声検出装置は、音声信号に伴い得る多様な騒音に柔軟に適応し、かつ性能の安定化が図られる。

追加開示請求の範囲3にかかわる音声検出装置では、品質監視手段12、16は、得られた音声信号の品質を順次積分し、その結果の単調増加関数あるいは単調非減少関数として得られる値をこの品質として適用する。

このような音声検出装置では、先行して与えられた音声フレームの品質が高く、あるいはその品質が高かった時間率が大きいほど、後続して与えられる音声フレームが有音区間である確度が大きな値で得られる。

#### 産業上の利用の可能性

本発明にかかわる第1、第2および第3の音声検出装置では、有音区間の内、例えば、子音区間のように音声信号の振幅が小さな領域に多く分布する区間については、その子音区間における音声信号の品質が低い場合であっても、有音区間であることを示す確度が大きな値で得られる。

本発明にかかわる第4の音声検出装置では、第1ないし第3の音声検出装置に比べて、音声フレーム毎に有音区間である確度が精度よく得られる。

本発明にかかわる第5の音声検出装置では、第1ないし第3の音声検出装置に比べて、処理量が削減され、あるいは応答性が高められる。

本発明にかかわる第6および第7の音声検出装置では、大きなレベルの雑音が重畳し、かつSN比が小さい音声フレームについても、有音区間に属することを示す確度が大きな値で得られる。



本発明にかかわる第1ないし第3の音声検出方法では、有音区間の内、例えば、子音区間のように音声信号の振幅が小さな領域に多く分布する区間については、その子音区間における音声信号の品質が低い場合であっても、有音区間であることを示す確度が大きな値で得られる。

したがって、これらの発明が適用された通信機器その他の電子機器では、音声信号を発する音響－電気変換手段が配置される音響的な環境、あるいはその音響信号の情報源の特性や性能に柔軟に適応しつつ、この音声信号の有音区間と無音区間との峻別が確度高く、安定に実現され、その峻別の結果に適応した所望の性能の達成と、資源の有効利用とが的確に図られる。

【図面の簡単な説明】

図1は、本発明の第1の原理ブロック図である。

図2は、本発明の第2の原理ブロック図である。

図3は、本発明の実施形態1、3～8を示す図である。

図4は、実施形態1の動作フローチャートである。

図5は、本発明の実施形態2を示す図である。

図6は、実施形態2の動作フローチャートである。

図7は、実施形態3の動作フローチャートである。

図8は、実施形態4の動作フローチャートである。

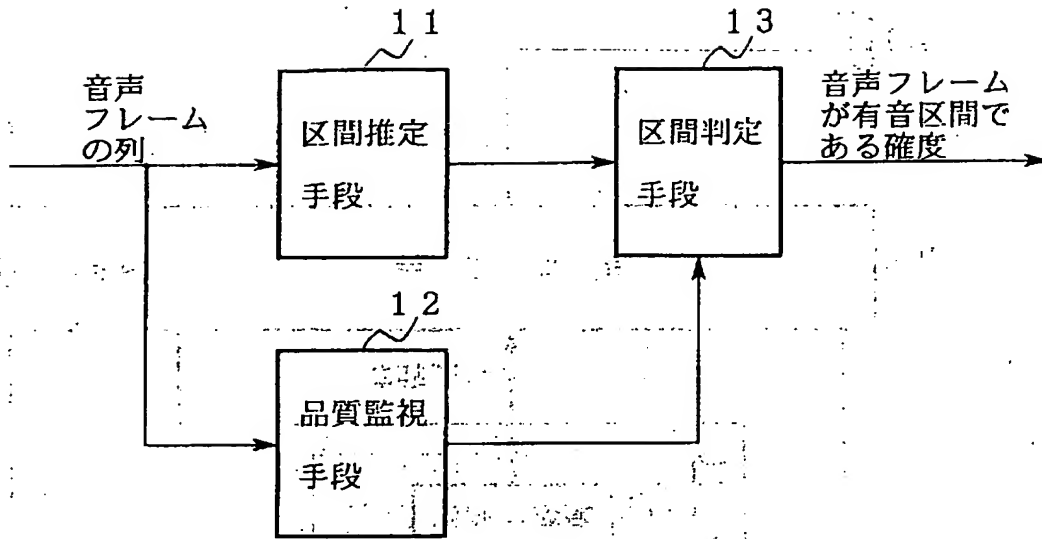
図9は、実施形態5の動作フローチャートである。

図10は、実施形態6の動作フローチャートである。

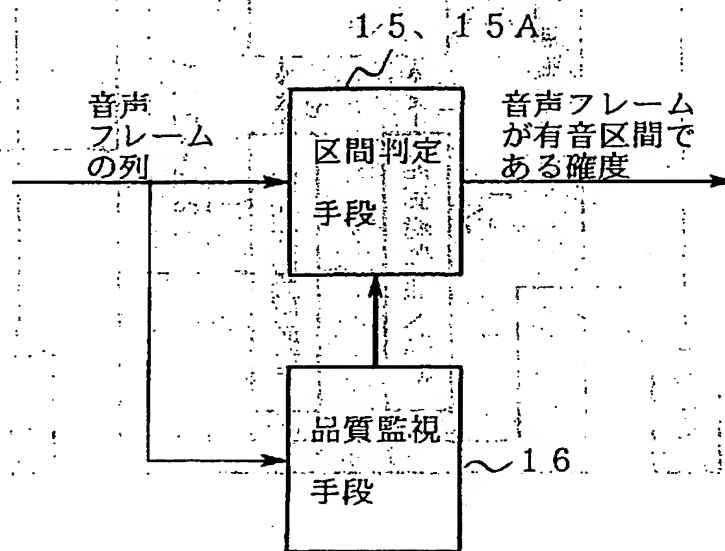
図11は、実施形態7および実施形態8の動作フローチャートである。

図12は、音声検出装置が搭載された無線端末装置の構成例を示す図である。

【図1】



【図2】



【図3】

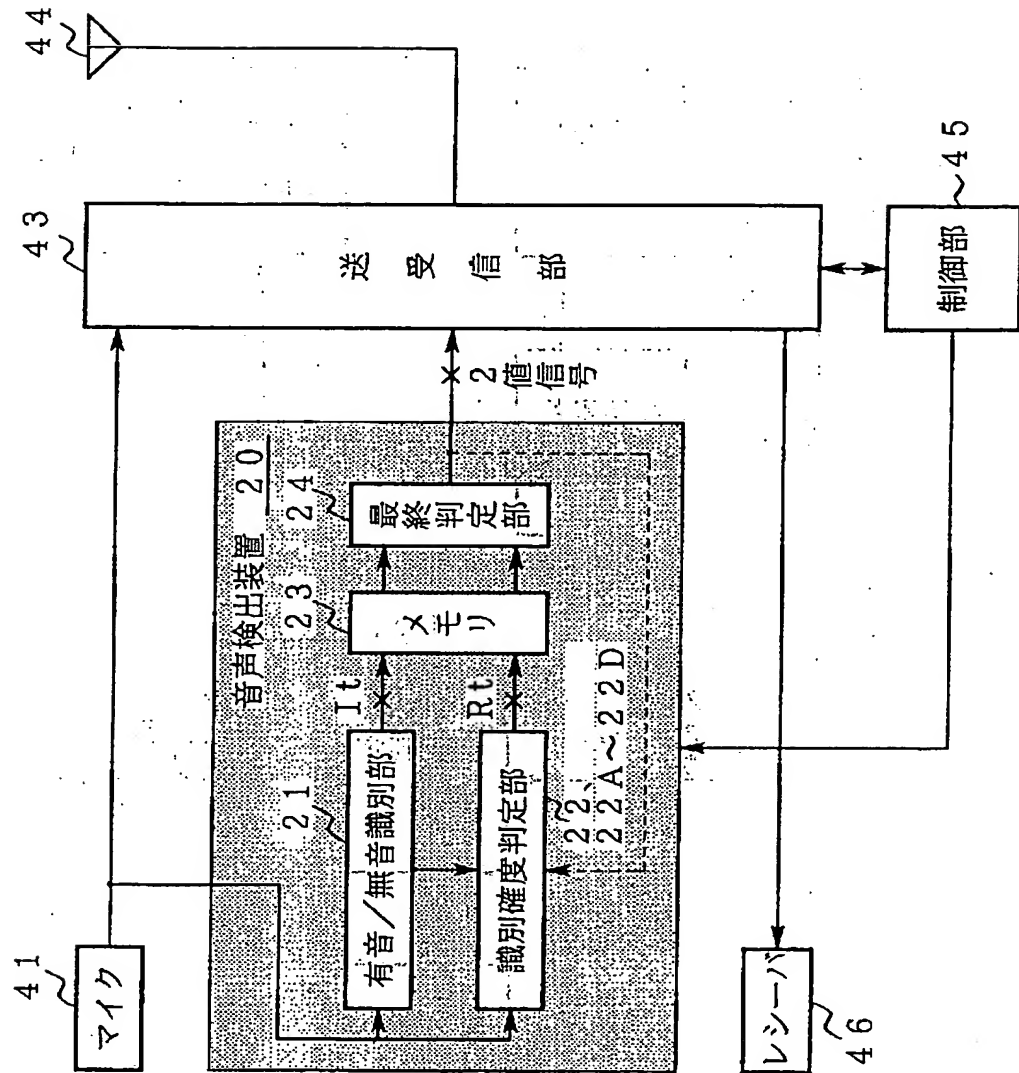
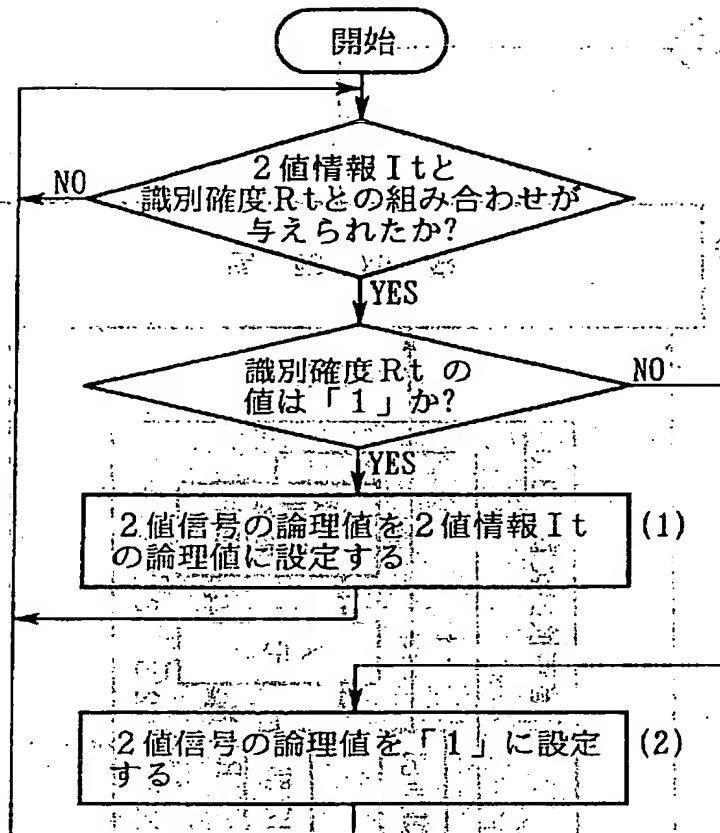


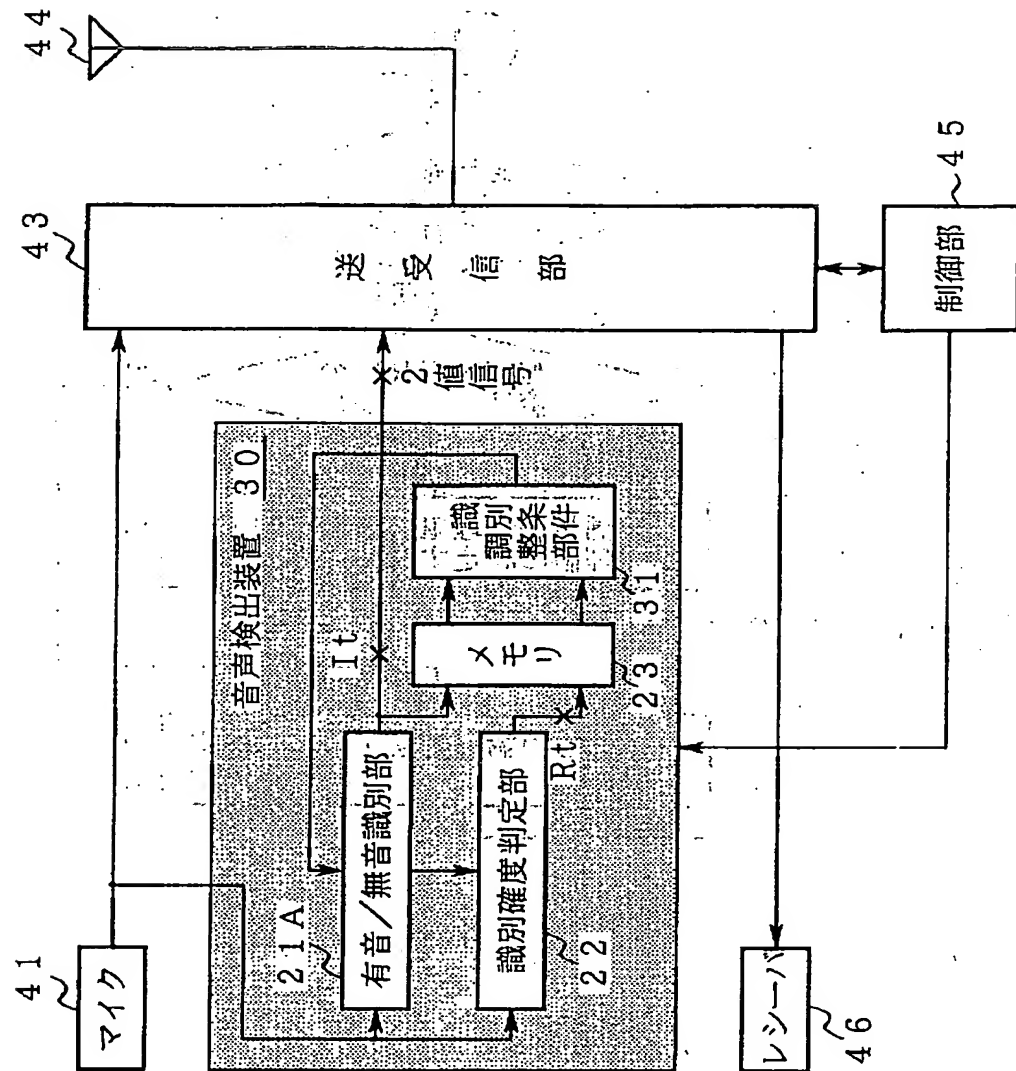
図3

【図4】

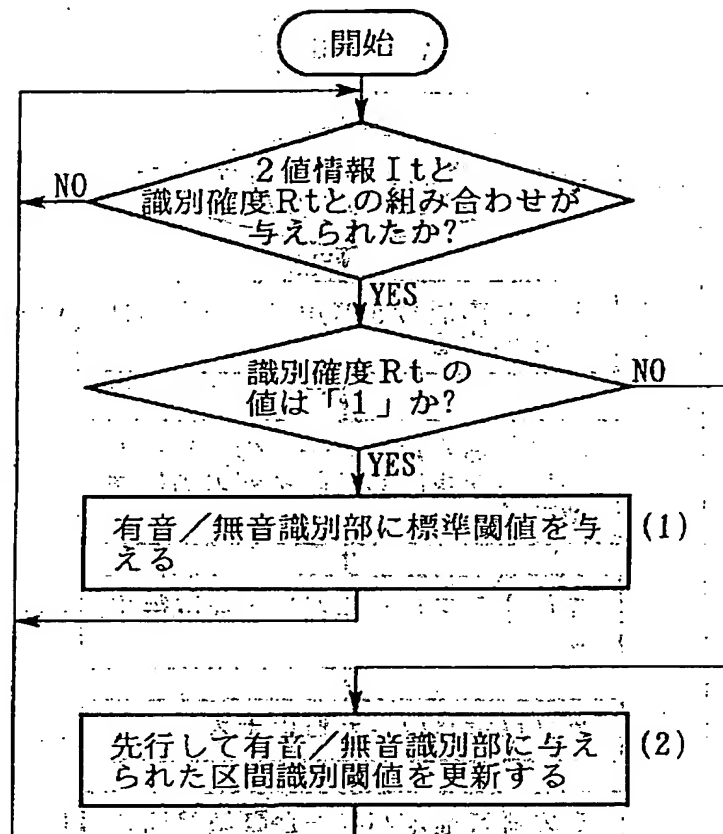


【図5】

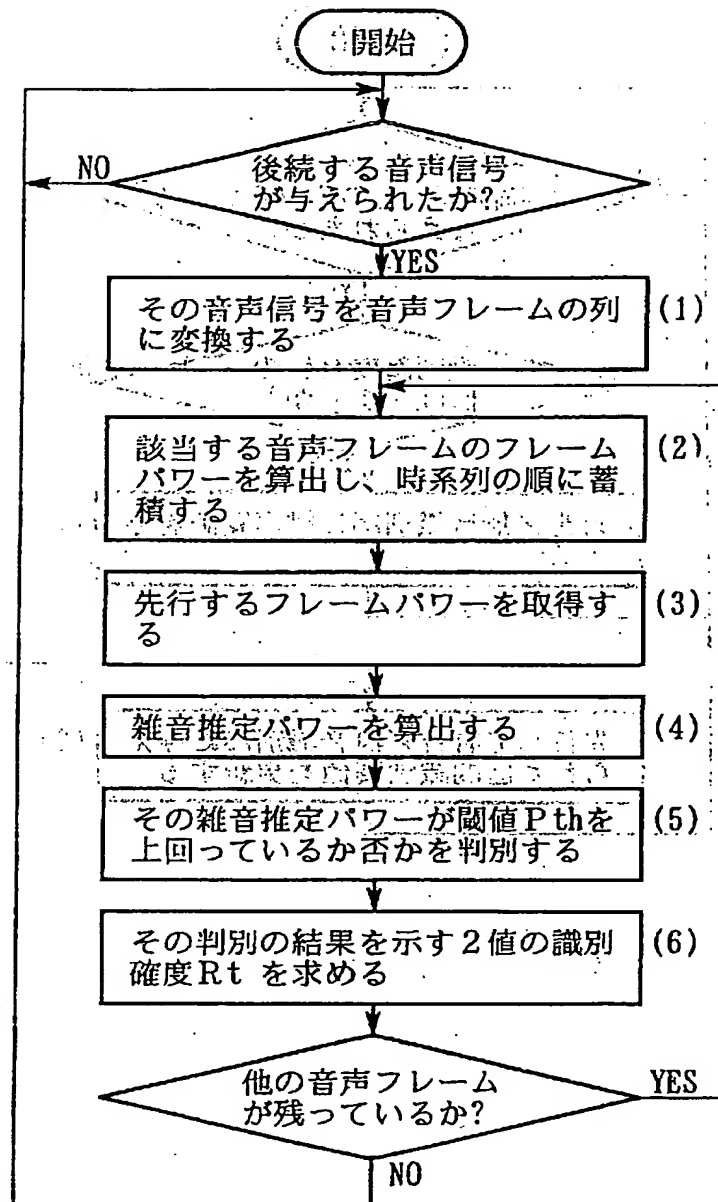
図5



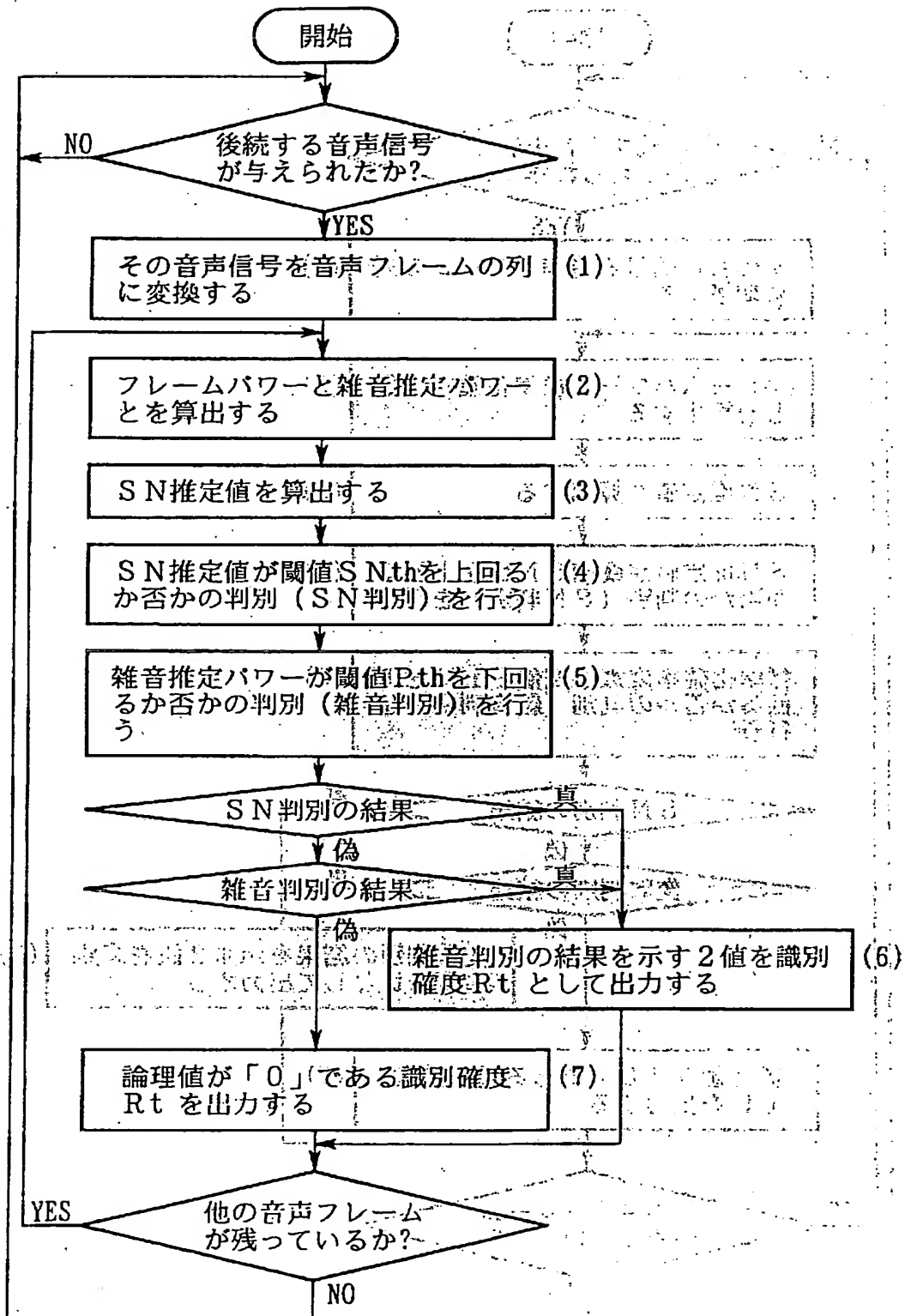
【図6】



【図7】

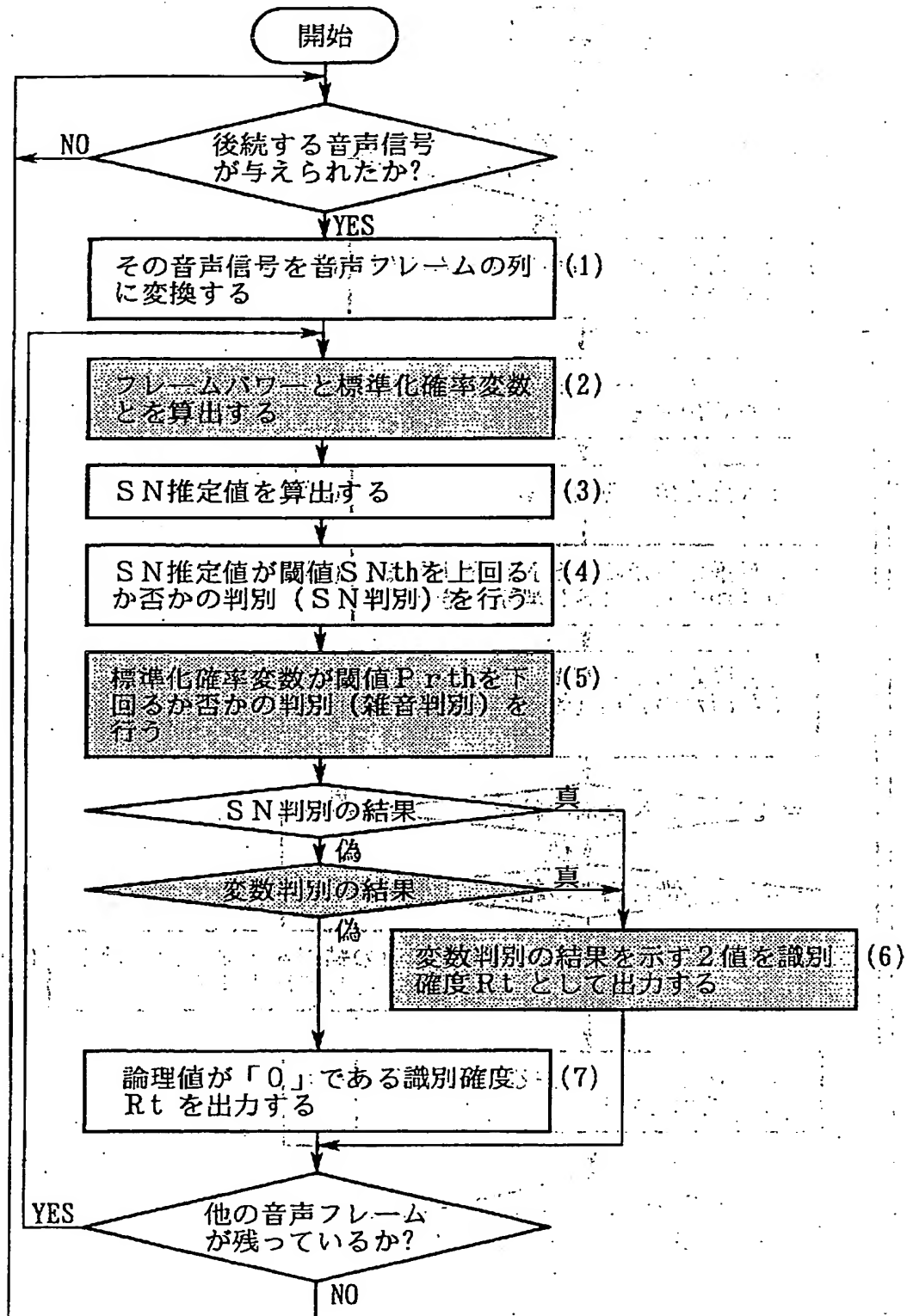


【図8】

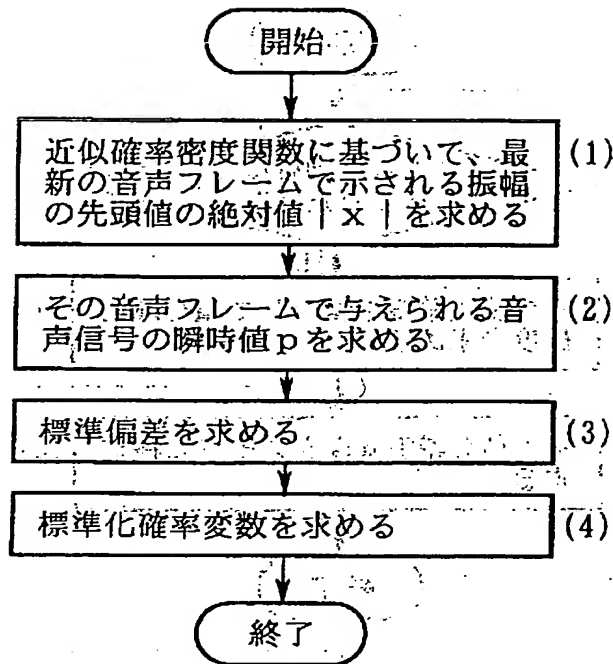




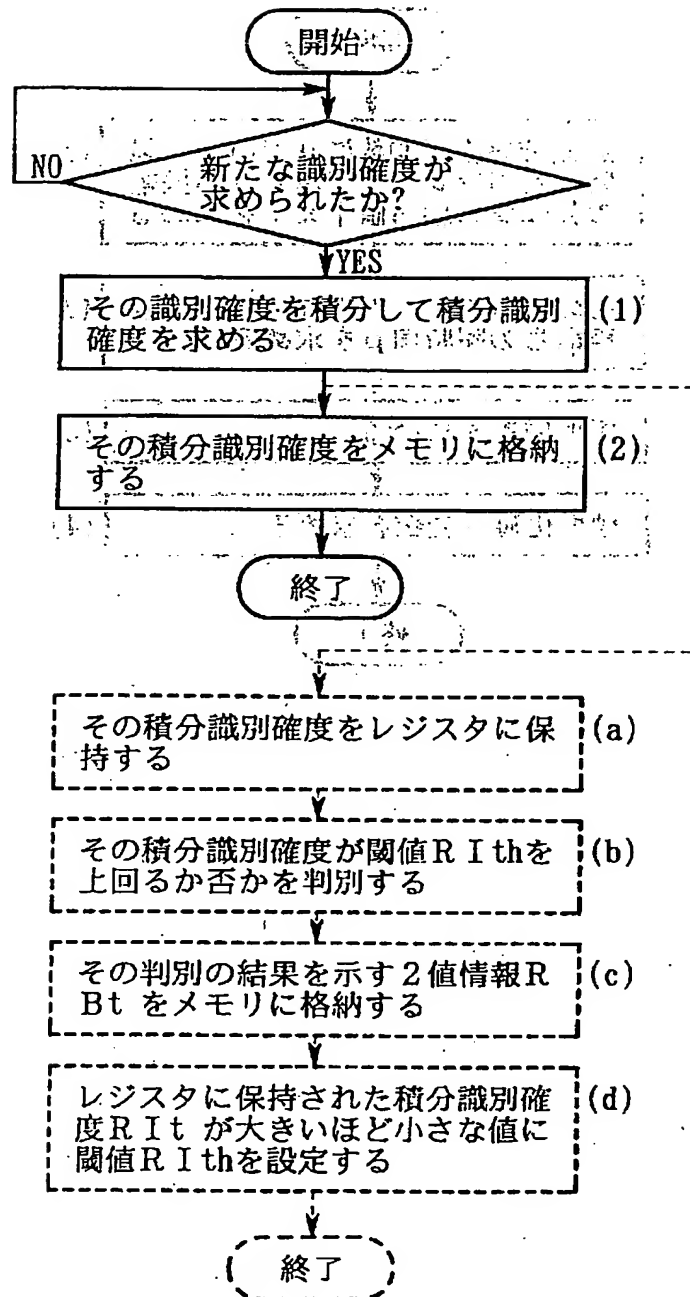
【図9】



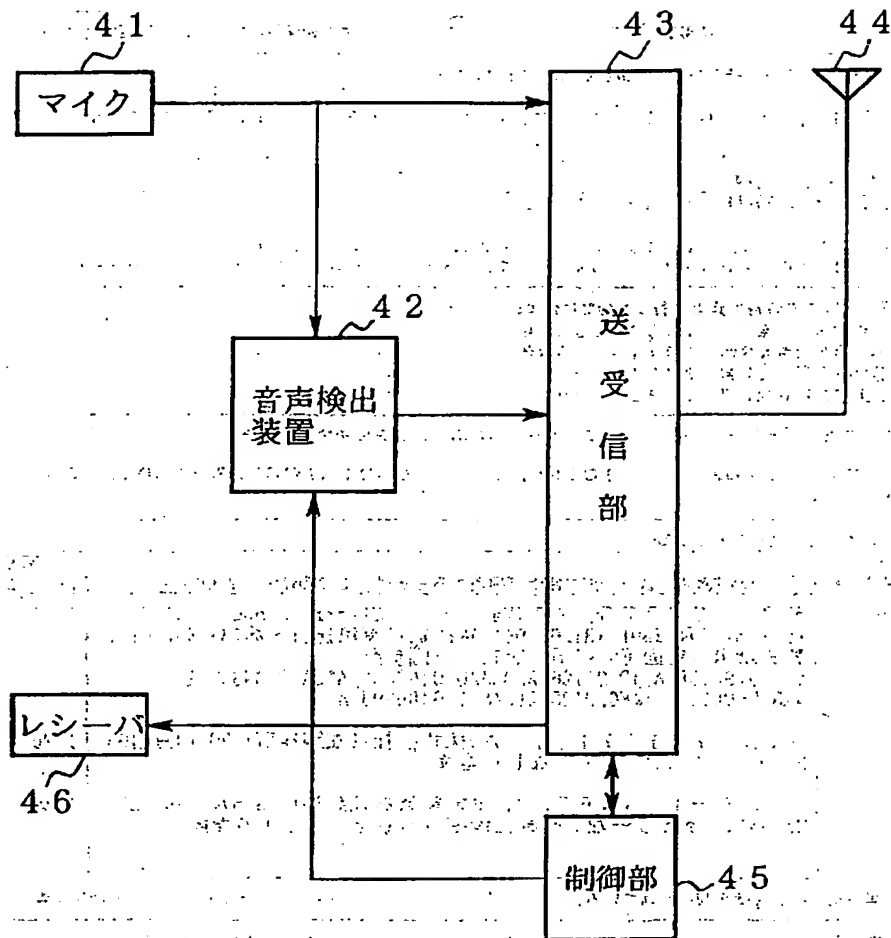
【図10】



【図11】



【図12】



## 【国際調査報告】

国際調査報告		国際出願番号 PCT/JP99/06639	
A. 発明の属する分野の分類 (国際特許分類 (IPC))			
Int Cl' G10L11/02, 21/02, H04B1/64, H04J3/17, H04Q7/32			
B. 調査を行った分野			
調査を行った最小限資料 (国際特許分類 (IPC))			
Int Cl' G10L11/02, 15/04, 15/20, 21/02, H04J3/17, H04Q7/00~38, H04B1/62~68, 3/00~60, H03G7/00~08			
最小限資料以外の資料で調査を行った分野に含まれるもの 日本国実用新案公報 1926~1995年 日本国公開実用新案公報 1971~1999年 日本国登録実用新案公報 1994~1999年 日本国実用新案登録公報 1996~1999年			
国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)			
JICST科学技術文献ファイル (JOIS), INSPEC (DIALOG), WPI (DIALOG), PATOLIS			
C. 関連すると認められる文献			
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号	
A	EP, 439073, A1 (Matsushita Electric Industrial Co. L td.) 31.7月:1991 (31.07.91) 特に第12頁第16行~第13頁第36行, 特許請求の範囲第8~9項; 第12~13図参照 &US, 5195138, A&JP, 2712692, B2&AU, 9168688, A&CA, 2034333, A &KR, 9606739, B1&NO, 9100221, A&FI, 9100293, A	1-20	
A	JP, 57-177197, A (株式会社日立製作所) 30.10月:1982 (30.10.82) (ファミリーなし) 全文	1-20	
A	JP, 5-130067, A (日本電気株式会社) 25.5月:1993 (25.05.93) (ファミリーなし) 特に段落0002~0010参照	1-20	
<input checked="" type="checkbox"/> C欄の続きにも文献が列挙されている。 <input type="checkbox"/> パテントファミリーに関する別紙を参照。			
* 引用文献のカテゴリー 「A」 特に関連のある文献ではなく、一般的技術水準を示すもの 「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの 「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す) 「O」 口頭による開示、使用、展示等に関する文献 「P」 国際出願日前で、かつ優先権の主張の基礎となる出願 の目後に公表された文献 「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの 「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの 「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの 「&」 同一パテントファミリー文献			
国際調査を完了した日		国際調査報告の発送日	
27.12.99		18.01.00	
国際調査機関の名称及びあて先 日本国特許庁 (ISA/JP) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号		特許庁審査官 (権限のある職員) 松尾 淳一 印 電話番号 03-3581-1101 内線 3540	

様式PCT/ISA/210 (第2ページ) (1998年7月)

## 国際調査報告

国際出願番号 PCT/JP99/06539

C (続き) 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	J P, 57-88497, A (株式会社日立製作所) 2.6月.1982 (02.06.82) (ファミリーなし) 全文	21
A	J P, 4-340598, A (株式会社リコー) 26.11月.1992 (26.11.92) (ファミリーなし) 全文	21

様式 PCT/ISA/210 (第2ページの続き) (1998年7月)

---

(注) この公表は、国際事務局 (W I P O) により国際公開された公報を基に作成したものである。

なおこの公表に係る日本語特許出願 (日本語実用新案登録出願) の国際公開の効果は、特許法第 184 条の 10 第 1 項 (実用新案法第 48 条の 13 第 2 項) により生ずるものであり、本掲載とは関係ありません。